Editorial

Dear Colleague:

Welcome to volume 24(5) of Intelligent Data Analysis (IDA) Journal.

This issue of the IDA journal is the fifth issue for our 24th year of publication. It contains thirteen articles representing a wide range of topics related to the theoretical and applied research in the field of Intelligent Data Analysis.

The first five articles are about data preprocessing and learning in IDA. In the first article Liu and Li argue that most existing hierarchical clustering algorithms are usually designed heuristically without an explicit objective function, which limits its utilization and analysis. The authors suggest combining Bayesian theory analysis with a K-means algorithm and introduce a hierarchical clustering based on K-means under the probability distribution framework. Their experimental results on both synthetic data and benchmark datasets demonstrate the effectiveness of their proposed algorithm over the existing related ones. Lei *et al*. in the second article of this issue explain that robust variable selection methods via penalized regression, such as least absolute deviation LASSO (LAD-LASSO), etc, have gained growing attention. However, those penalized regression procedures are still sensitive to noisy data. Focusing on the shrinkage estimation and variable selection tasks on noisy streaming data, the authors present a noise-resilient online learning regression model which is resistant to noisy data in both explanatory variables and response variables. Their extensive simulation studies demonstrate satisfactory sparseness and noise-resilient performances. Garcia *et al*. in the next article argue that Meta-Learning has been largely used over the last few years to support the recommendation of the most suitable machine learning algorithm(s) and hyperparameters. However, if one wants the use of Meta-Learning to be computationally efficient, the extraction of the meta-feature values should also show a low computational cost, considering a trade-off between the time spent to run all the algorithms and the time required to extract the meta-features. The authors propose an empirical approach designed to decrease the computational cost of computing the data complexity measures, while still keeping their descriptive ability and show that the predicted data complexity measures is similar to the performance obtained using the original data complexity measures. Lonlac and Nguifo in the next article of this issue explain that for mining frequent simultaneous attribute co-variations in numerical databases few efficient algorithms for automatically extracting such patterns have been reported in the literature. The authors propose an approach that considerably reduces the number of gradual patterns within an ordered data set. The experimental results show the benefits of their approach. Fize *et al*. in the last article of this group discuss that textual data is available to an increasing extent through different media and new information extraction methods are needed since these new resources are highly heterogeneous. The authors propose a text matching process based on spatial features and assessed through heterogeneous textual data and propose an approach with new contributions such as: a new geocoding methods, a thorough evaluation and an in-depth discussion. Their results obtained on two corpora demonstrate good spatial matches can be obtained between the most similar spatial textual representations.

The second group of articles in this issue are about neural nets in IDA. Zvarevashe and Olugbara in the first article of this group introduce a custom 2D-convolution neural network that performs both feature extraction and classification of vocal utterances in speech emotion recognition. The authors evaluate their

neural network against deep multilayer perceptron neural network and deep radial basis function using the Berlin database of emotional speech. They conclude that there may be a need to develop customized solutions for different language settings depending on the area of applications. Guo and Li in the seventh article of this issue propose a hybrid algorithm with the strategy of two-stage searches in which it firstly determines the local search space based on Maximal Information Coefficient and then searches the local space by Binary Particle Swarm Optimization. In the second-stage of their proposed approach, an efficient algorithm based on three basic operators is proposed to extend the local space to the whole space. Their experiment results show that the proposed algorithm can obtain better performance of BN structure learning. Uteuliyeva *et al.* in the next article provide a review of neural network architectures which were motivated by Fourier series and integrals and which are referred to as Fourier neural networks. These networks are empirically evaluated in synthetic and real-world tasks. Neither of them outperforms the standard neural network with sigmoid activation function in the real-world tasks. The authors conclude that all neural networks, both Fourier and the standard one, empirically demonstrate lower approximation error than the truncated Fourier series when it comes to approximation of a known function of multiple variables.

And finally the third group of articles are about enabling techniques and innovative application in IDA. Borges *et al.* in the first article of this group explain that discovering motifs in time series data has been widely explored and various techniques have been developed to tackle this problem. However, when it comes to spatial-time series, a clear gap can be observed. The authors address this problem by presenting an approach to discover and rank motifs in spatial-time series, through denominated Combined Series Approach. In their approach, motifs are validated according to both temporal and spatial constraints and are ranked according to their entropy, the number of occurrences, and the proximity of their occurrences. The approach is evaluated using both synthetic and real datasets. The next article by Sepúlveda and Norambuena is about Twitter Sentiment Analysis for the estimation of voting intention in Chilean elections. The idea is to estimate the voting intention associated with each candidate in order to contrast this with the results from classical methods (e.g., polls and surveys). The authors acquired the data, labelled the tweets as positive or negative and built a model using machine learning techniques. Their classification model had a relatively high accuracy using support vector machines, which yielded the best model for their case. Tai and Hsu in the next article of this group discuss profitability and errors of predicted prices from deep learning via program trading. The authors propose to find the parameter sets of their models with low magnitude-based error and then use program trading to find out their profitability. Their results indicate that, in assessing the performance of deep learning, how to use the predicted values in applications and the application results could also be part of the quality measurement for the model assessment in the learning. Huang *et al.* in the twelfth article of this issue introduce autonomous self-evolving forecasting models for price movement in high frequency trading (HFT). The authors argue that forecasting financial time series data has been a challenging task because this kind of data is typically quite noisy and non-stationary. They develop novel computational intelligence based methodologies for the forecasting task of price movement in HFT with the goal of conducting a study for autonomous genetic-based models that allow the forecasting systems to self-evolve. Their results show that their proposed method can improve upon the previous ones and advance the current state of research. And finally the last article of this issue by Pham and Do is about heterogeneous information networks (HIN) where the authors argue that information network embedding has become an effective approach for information network analysis and mining tasks. The authors introduce a novel approach of topic-driven meta-path-based embedding, in which discovered models enable to capture richer semantic of node representation by applying the meta-path-based community-aware, node proximity preserving and topic

similarity evaluation at the same time as the process of network embedding. The authors demonstrate comprehensive empirical studies of their proposed approach with several real-world HINs.

In conclusion, we would like to thank all the authors who have submitted the results of their excellent research to be evaluated by our referees and published in the IDA journal. This year, we have prepared a special issue which is from ten best papers presented during the CIARP-2019 conference that was held in Havana-Cuba, in October 2019. This year's special issue will be published before the last issue of this year. We look forward to receiving your feedback along with more and more quality articles in both applied and theoretical research related to the field of IDA.

With our best wishes,
Dr. A. Famili
Editor-in-Chief